

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Y. Wakita et al. : Art Unit:
Serial No.: To Be Assigned : Examiner:
Filed: Herewith :
FOR: METHOD AND APPARATUS FOR :
CONVERTING EXPRESSION :
: 

CLAIM TO RIGHT OF PRIORITY

Assistant Commissioner for Patents
Washington, D.C. 20231
SIR:

Pursuant to 35 U.S.C. 119, Applicants' claim to the benefit of filing of prior Japanese Patent Application No. 2000-066494, filed March 10, 2000, is hereby confirmed.

A certified copy of the above-referenced application is enclosed.

~~Respectfully Submitted,~~

~~Allan Ratner, Reg. No. 19,717~~
Attorney for Applicants

AR/dlm

Enclosure: (1) priority document

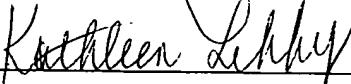
Dated: March 12, 2001

Suite 301
One Westlakes, Berwyn
P.O. Box 980
Valley Forge, PA 19482-0980
(610) 407-0700

The Assistant Commissioner for Patents is
hereby authorized to charge payment to
Deposit Account No. 18-0350 of any fees
associated with this communication.

EXPRESS MAIL Mailing Label Number: EL 736966386 US
Date of Deposit: March 12, 2001

I hereby certify that this paper and fee are being deposited, under 37 C.F.R. § 1.10 and with sufficient postage, using the "Express Mail Post Office to Addressee" service of the United States Postal Service on the date indicated above and that the deposit is addressed to the Assistant Commissioner for Patents, Washington, D.C. 20231.


Kathleen Libby

日本国特許庁
PATENT OFFICE
JAPANESE GOVERNMENT

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出願年月日

Date of Application:

2000年 3月10日

出願番号

Application Number:

特願2000-066494

出願人

Applicant (s):

松下電器産業株式会社

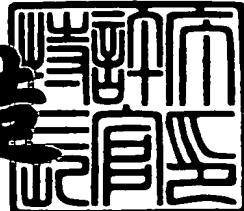
03/12/01
09/803779 PRO
11000 U.S.



2001年 2月23日

特許庁長官
Commissioner,
Patent Office

及川耕造



出証番号 出証特2001-3011032

【書類名】 特許願
【整理番号】 2033820024
【提出日】 平成12年 3月10日
【あて先】 特許庁長官殿
【国際特許分類】 G10L 3/00
【発明者】
【住所又は居所】 大阪府門真市大字門真1006番地 松下電器産業株式
会社内
【氏名】 脇田 由実
【発明者】
【住所又は居所】 大阪府門真市大字門真1006番地 松下電器産業株式
会社内
【氏名】 松井 謙二
【特許出願人】
【識別番号】 000005821
【氏名又は名称】 松下電器産業株式会社
【代理人】
【識別番号】 100097445
【弁理士】
【氏名又は名称】 岩橋 文雄
【選任した代理人】
【識別番号】 100103355
【弁理士】
【氏名又は名称】 坂口 智康
【選任した代理人】
【識別番号】 100109667
【弁理士】
【氏名又は名称】 内藤 浩樹

【手数料の表示】

【予納台帳番号】 011305

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9809938

【書類名】 明細書

【発明の名称】 表現変換方法

【特許請求の範囲】

【請求項 1】 ある入力文を、前記入力文が持つ意図あるいは文意を理解するための単語（以下、重要語という）を用いて、同じ意図あるいは文意を表す入力文とは別の表現文に変換する表現変換方法。

【請求項 2】 別の表現文は標準的または簡易的表現文であり、入力文に含まれる一部の内容語の単語もしくはその変換された単語を重要語とし、この重要語と入力文から推定される文意により決まる表現との組み合わせからなることを特徴とする請求項 1 記載の表現変換方法。

【請求項 3】 文意の推定は、入力文から予め定められた重要語のみを抽出し、各重要語の共起関係または依存関係を用いて推定することを特徴とする請求項 2 記載の表現変換方法。

【請求項 4】 文意の推定は、予め、出力される標準的または簡易的表現パターンとそのパターンに当たはまる重要語群及び重要語群間の共起関係または依存関係とを対応させている用例文データベースを作成しておき、入力文から重要語群を抽出し、抽出された重要語群と最も類似している重要語群をもつ表現パターンを用例文データベースから選択することであることを特徴とする請求項 2 に記載の表現変換方法。

【請求項 5】 用例文データベースの表現パターンは、重要語または重要語の対訳単語のみで構成されていることを特徴とする請求項 4 記載の表現変換方法。

【請求項 6】 用例文データベースの表現パターン及び重要語群は、重要語を含む単語クラスとして記述されていることを特徴とする請求項 4 記載の表現変換方法。

【請求項 7】 文意の推定は、入力文から予め定められた重要語のみを抽出し、各重要語間の関係から入力誤りである単語を推定し、推定された単語を除いた残りの重要語から文意を推定し、推定される文意により決まる単語の組み合わせからなることを特徴とする請求項 2 記載の表現変換方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、翻訳または通訳などの言語変換手段、または書き言葉を話し言葉に変換する文型変換手段、または、複雑または冗長な文章を要約して出力する要約文作成手段などに関するものである。

【0002】

【従来の技術】

以下、従来技術を通訳ソフトを例にして説明する。

【0003】

通訳ソフトは、音響信号として入力された発声文を単語テキスト列で表示された出力文に変換するための音声認識と、単語テキスト列で表示された文を入力し他言語文に翻訳する表現変換としての言語翻訳とを順次実行することで通訳を実現している。さらに上記言語翻訳部は、入力文の統語的または意味的構造を解析する言語解析部と、解析結果に基づいて他言語に変換する言語変換部と、翻訳結果から自然な出力文を生成する出力文生成部とから構成されている。通訳ソフトや一部の翻訳ソフトでは、話し言葉に頻繁に見られるような、文法規則から外れたフランクな表現を正しく翻訳するために、文法規則に従った言語解析を行うのではなく、実際に入力される文と類似した用例文を学習しておき、類似用例を検索することで言語解析を行う手法を使用している場合が多い。以下、このような従来の通訳ソフト例を図3の例に沿って説明する（たとえば、古瀬、隅田、飯田：情報処理学会論文誌Vol35,no3,1994-3）。

【0004】

通訳を行う前に、予め発声文対訳用例コーパスから発声文規則を抽出する。ここでは、一部の単語が認識誤りを起したり省略して発声されることを考慮し、発声文例を意味的なまとまりとしての最小単位（以後、フレーズと呼ぶ）に分割し、フレーズ内規則とフレーズ間の依存規則とを作成している。まずフレーズ決定部61で、対訳発声文例はフレーズ単位に分割され、対訳フレーズ辞書作成部62では、対応フレーズ辞書が、フレーズ内の内容語を変数化された形で作成される。たとえば、「部屋の予約をお願いしたいんですが」「I'd like to reserve a

room」という対訳用例は、(a)「部屋の予約」「reserve a room」,(b)「お願いしたいんですが」「I'd like to」という(a)(b)2つの対訳フレーズに分割される。そして「部屋」「予約」「お願い」などの内容語は各々XYZという変数で表され、(a)「XのY」「YX」、(b)「Zしたいんですが」「I'd like to」という2つの対訳フレーズ規則が対訳フレーズ辞書63に記述される。また、各々のフレーズの順序関係を規則化するため、「(a)を(b)する」「(b) to (a)」というフレーズ間の関係が、フレーズ間規則テーブル64に各々保管される。このような処理が対訳コーパスに含まれた全発声文分について行われる。

【0005】

通訳の際には、まず発声された原言語音声は音声認識部64に入力される。音声認識部では、たとえば、対訳フレーズ辞書にフレーズとして記述されている単語列とフレーズ間規則にて記述されているフレーズ列とから予測される可能性のある単語列の中から、音響的に最も類似している単語列を音声認識候補として出力する。認識された連続単語列は、言語翻訳部65に入力される。言語翻訳部では、入力された連続単語列をフレーズ辞書内のフレーム列に変換し、各フレーム列に相当するフレーズ間規則を探索する。そして、各フレーズの対訳である目的言語フレーズと目的言語のフレーズ間規則とから、入力原言語認識結果文を目的言語文に変換する。

【0006】

変換された目的言語文は出力文生成部66に入力され、文法的な不自然さを修正する。たとえば、代名詞や動詞や助動詞における3人称化や複数化や過去形化などの最適化や、文全体の構成の最適化などが行われる。修正後の目的言語翻訳結果文はたとえばテキストとして出力される。

【0007】

【発明が解決しようとする課題】

しかしながら、従来手段では、文法規則から外れた入力文に対応できる良さがある反面、多数の異なる対訳フレーズとその組み合わせが、そのまま規則として記述されるため、変換規則が煩雑で膨大になってしまい処理時間もかかってしまう、という課題を有している。また、出力文生成部にて行われる文法的なチェック

クにも、複雑な規則が必要となり、特にゼロ代名詞の補完などは完全に正しく補完できる技術がなく、時々は誤った補完を行ってしまうという課題を有している。また、音声認識誤りなどにより、部分的に誤っている文が言語翻訳部に入力された場合、誤ったフレーズ内およびフレーズ間規則に基づき言語変換がなさるため、全く意図が伝わらない翻訳結果を出力するという課題を有している。

【0008】

本発明の目的は、以上の課題を解決し、入力文中から、重要語を選択し、選択された単語のみから表現変換文にあたる目的言語文を構成することで、言語変換規則と出力文生成規則の簡略化をはかることで、コンパクトで高速に処理可能な表現変換手段を提供するものである。また、重要語のみを用いて意図推定と言語変換処理を行うことにより、重要語以外で認識誤りを起こしても、出力文の質に悪影響を及ぼすことなく意図が正しく伝わる結果を出力できる。また、選択された重要語間の共起関係や依存関係を用いて認識誤りを起こした単語を推定し、認識誤り部分を除去してから文意を推定することにより、従来のように全く文意が伝わらない結果を出力してしまう、という課題を回避することができる表現変換手段を提供するものである。

【0009】

【課題を解決するための手段】

上述した課題を解決するために、請求項1記載の表現変換手段は、入力文から重要語を抽出する重要語抽出手段と、重要語から文意を理解する意図分析手段と、重要語を用いて標準的または簡易的表現文に変換し出力する言語変換手段とから構成される。

【0010】

上述した課題を解決するために、請求項2に記載の表現変換手段は、入力文に含まれる一部の内容語の単語もしくはその変換された単語を重要語として抽出する重要語抽出手段と、重要語から文意を推定する意図推定手段と、重要語と入力文から推定される文意により決まる表現との組み合わせからなる標準的または簡易的表現を生成する言語変換手段から構成される。

【0011】

上述した課題を解決するために、請求項3に記載の表現変換手段は、入力文から重要語を抽出する重要語抽出手段と、各重要語の共起関係または依存関係から意図を推定する意図分析手段と、重要語と入力文から推定される文意により決まる表現との組み合わせからなる標準的または簡易的表現を生成する言語変換手段から構成される。

【0012】

上述した課題を解決するために、請求項4に記載の表現変換手段は、出力される標準的または簡易的表現パターンとそのパターンに当てはまる重要語群及び重要語群間の共起関係または依存関係とを対応させている用例文データベースと、入力文から重要語を抽出する重要語抽出手段と、抽出された重要語群と最も類似している重要語群をもつ表現パターンを用例文データベースから選択する意図分析手段と、選択された用例文を出力する言語変換手段とから構成される。

【0013】

上述した課題を解決するために、請求項5に記載の表現変換手段は、入力文から重要語を抽出する重要語抽出手段と、重要語から意図を推定する意図分析手段と、重要語または重要語の対訳単語のみから標準的または簡易的表現を生成する言語変換手段から構成される。

【0014】

上述した課題を解決するために、請求項6に記載の表現変換手段は、重要語が重要語を含む単語クラスで記述された標準的または簡易的表現パターンと各パターンに当てはまる重要語クラス群及び重要語クラス群間の共起関係または依存関係とを対応させている用例文データベースと、入力文から重要語を抽出する重要語抽出手段と、抽出された重要語を重要語クラス記述にかかる単語クラス化手段と、抽出された重要語クラス群と最も類似している重要語クラス群をもつ表現パターンを用例文データベースから選択する意図分析手段と、選択された用例文の重要語クラスを重要語に変換して出力する言語変換手段とから構成される。

【0015】

上述した課題を解決するために、請求項7に記載の表現変換手段は、入力文から重要語を抽出する重要語抽出手段と、各重要語の共起関係または依存関係から

入力誤り単語を推定する誤認識単語推定手段と、推定された単語を除いた残りの重要語から意図を推定する意図分析手段と、重要語と入力文から推定される文意により決まる表現との組み合わせからなる標準的または簡易的表現を生成する言語変換手段から構成される。

【0016】

次に本発明の動作を説明する。

【0017】

請求項1に記載の表現変換手段においては、入力文から重要語を抽出し、抽出された重要語を用いて同じ文意を表す標準的または簡易的表現文に変換し出力することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0018】

請求項2に記載の表現変換手段においては、入力文に含まれる一部の内容語の単語もしくはその変換された単語を重要語として抽出し、この重要語と入力文から推定される文意により決まる表現との組み合わせからなる標準的または簡易的表現文を生成することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0019】

請求項3に記載の表現変換手段においては、入力文から重要語を抽出し、抽出された重要語の共起関係または依存関係から文意を推定し、重要語またはその対訳単語と推定された意図から予め決められている単語のみの組み合わせから標準的または簡易的表現を生成することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表

現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0020】

請求項4に記載の表現変換手段においては、予め、出力される標準的または簡易的表現パターンとそのパターンに当たる重要な語群及び重要語群間の共起関係または依存関係とを対応させている用例文データベースを作成しておき、入力文から重要な語群を抽出し、抽出された重要な語群と最も類似している重要な語群をもつ表現パターンを用例文データベースから選択し、選択された用例文を出力することにより、請求項1, 2, 3に記載の表現変換手段の効果に加え、現実に入力される文の種類やドメインや文型に忠実に確度よく表現変換を行うことを可能にする。

【0021】

請求項5に記載の表現変換手段においては、請求項4に記載の用例文データベースに記載されている表現パターンが重要な語または重要な語の対訳単語のみで構成されていることにより、請求項1, 2, 3, 4に記載の効果をさらに強化することが可能である。

【0022】

請求項6に記載の表現変換手段においては、請求項4に記載の用例文データベースに記載されている表現パターン、重要な語群及び重要な語群間の共起関係または依存関係の重要な語の記述を重要な語を含む単語クラスの記述として、請求項1, 2, 3, 4に記載の効果に加え、用例文データベースに含まれていない重要な語が入力されても妥当な用例文を選択可能となり、より多種多様な入力文に対応できる表現変換が可能となる。

【0023】

請求項7に記載の表現変換手段においては、入力文から重要な語群を抽出し、抽出された重要な語間の関係から入力誤りである単語を推定し、推定された単語を除いた残りの重要な語から文意を推定し、推定される文意により決まる単語の組み合わせから標準的または簡易的表現を生成することで、請求項3に記載した効果に加えて、たとえ重要な語が誤っていても、その誤りの重度に応じて、正しく文意が

理解できる表現に変換するか、意味が理解できなかったことをユーザに告知することが可能となり、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0024】

【発明の実施の形態】

以下、図面を参照して本発明の請求項1から5の実施例を説明する。本実施例では、表現変換手段の一例として、従来例同様、音声にて入力された原言語入力文を他言語である目的言語文に変換する通訳手段を用いて説明する。図1は本発明の請求項1から5に係る一実施例である通訳手段の構成図である。

【0025】

実施例の通訳手段は、予め通訳する前に、重要語の内容と重要語群と目的言語表現パターンとの対応を決定し、重要語対訳辞書、変換用用例DBに記載しておく必要がある。これらは、開発者が、入力文意毎に意図を表す重要語及びそれを用いた表現パターンを手動で決めてよいし、対訳コーパスの各対訳文毎に意図タグを付け、意図毎に対訳コーパスを分類し、各文意に共通に使用される単語を重要語候補と決め、それらを開発者がチェックすることで半自動的に重要語と表現パターンを決定してもよい。上記のような手段で決められた重要語と表現パターンから、重要語対訳辞書、変換用用例DBを作成しておく（図2（a））。

【0026】

ただし、変換用用例DBには、以下の手順で重要語の依存関係を付加する。まず対訳コーパスの原言語文に対して、依存構造分析を行い、各フレーズ毎の依存構造を明らかにする。この依存構造関係で依存関係にある重要語対があれば、その情報を、相当する用例DB重要語と表現パターン対に追加する。各々の例を図2に示す。このタグ付きコーパスからは、例にある重要語と用例DBが作成され、各重要語の依存関係が付加されている。

【0027】

通訳の際には、原言語音声が入力され、音声認識部で認識結果である単語列候補が出力される。この認識結果は、重要語抽出部に入力される。重要語抽出部では予め決められた重要語が認識結果文から抽出される。次に意図理解部に用例文

選択部で、入力文中の重要語と用例文中の重要語対とを比較し、入力文中に最も多く重要語対が含まれている用例文を、最も入力文と文意が類似した用例文であるとみなして選択し、出力する。

【0028】

以上の実施例では、予め、出力される目的言語文の標準的または簡易的表現パターンとそのパターンに当てはまる重要語群及び重要語群間の依存関係とを対応させている用例文データベースを作成しておき、入力文から重要語し、抽出された重要語群と最も類似している重要語群をもつ表現パターンを用例文データベースから選択し出力することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。

【0029】

また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0030】

なお、請求項5に示した例のように、予め作成された用例文データベースに記述された表現パタンは図3に示したように重要語のみで構成されている場合には、なお、変換規則と文生成規則がコンパクトに構成する事が可能となり、効果の高い表現生成手段が可能となる。

【0031】

次に請求項6における実施例を説明する。実施例の通訳手段は、前の実施例と同様に、予め通訳する前に、重要語の内容と重要語群と目的言語表現パタンとの対応を決定し、重要語辞書、重要語対訳辞書、変換用用例DBに記載しておき、依存構造関係で依存関係にある重要語対も同様に表現パタン対に追加する。さらに、分類語意表の意味コードに従って、用例データベースにある重要語、依存関係をなしている重要語対、変換規則に含まれる重要語を意味コードに置き換えて、各重要語をクラス化しておく。クラス化された重要語対訳辞書と用例文データベースの例を図5に示す。図5の分類語彙表では「コーヒー」などの飲み物を示す単語の意味コードが100であり、「冷たい」などの飲み物の状態を表す意味コ

ードが101である。各々の意味コードを用いて、先の図2の用例DBの中の重要語は意味コードで表されている。

【0032】

通訳の際には、原言語音声が入力され、音声認識部で認識結果である単語列候補が出力される。この認識結果は、重要語抽出部に入力される。重要語抽出部では予め決められた重要語が認識結果文から抽出される。分類語彙表の意味コードに従って重要語が意味コードに置き換えることで、重要語を単語クラス化する。次に意図理解部に相当する用例文選択部で、入力文中の重要語クラスと用例文中の重要語クラス対とを比較し、入力文中に最も多く重要語クラス対が含まれている用例文を、最も入力文と文意が類似した用例文であるとみなして選択する。出力生成部では、選択した用例文の重要語クラスを同じクラスに含まれる入力された重要語に戻し出力する。

【0033】

以上の実施例では、予め、出力される目的言語文の標準的または簡易的表現パターンとそのパターンに当てはまる重要語クラス群及び重要語クラス群間の依存関係とを対応させている用例文データベースを作成しておき、入力文から重要語を抽出し、抽出された重要語重要語を分類語彙表に従ってクラス化し、クラス化された重要語群と最も類似している重要語クラス群をもつ表現パターンを用例文データベースから選択し、クラス化された重要語をもとの重要語に戻して出力することで、前の実施例に示した効果のように変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。さらに、単語クラス化することで、用例文にはない新たな入力文に対しても正しい翻訳結果を出力することが可能となるため、少ない用例文データベースで多くの入力文に対応できる表現変換手段を可能とする。

【0034】

次に請求項7における実施例を説明する。本実施例も前実施例同様、通訳手段

を例に説明する。実施例の通訳手段は、予め通訳する前に、重要語の内容と重要語群と目的言語表現パターンとの対応を決定し、重要語辞書、重要語対訳辞書、変換用用例DBに記載し、重要語の依存関係も同様に変換用用例データベースに付加する。

【0035】

通訳の際には、原言語音声が入力され、音声認識部で認識結果である単語列候補が出力される。この認識結果は重要語抽出部に入力され、予め決められた重要語が認識結果文から抽出される。次に誤認識単語抽出部で、用例文データベースにある重要語対と抽出された重要語群とを比較し、抽出された重要語の中で、他の重要語のどれとも依存関係による重要語対にならない重要語を誤認識単語と推定する。次に用例文検索部で、先の誤認識単語推定部で誤認識であるとみなされた単語を除いた残りの重要語を用いて、前の実施例と同様に最も入力文と意図が類似した用例文を選択し出力する。図7に例を示す。「熱いミルク」と入力されたにもかかわらず、「青いミルク」と誤認識を尾もした例である。

【0036】

用例DBの依存関係と比べた場合、「青い」と依存関係を持つ単語が認識結果の中に存在しないため、「青い」を誤認識単語であると推定し、「ミルクはありますか」という認識結果として用例文を検索する。

【0037】

以上の実施例では、予め、出力される目的言語文の標準的または簡易的表現パターンとそのパターンに当てはまる重要語群及び重要語群間の依存関係とを対応させている用例文データベースを作成しておき、入力文から重要語し、抽出された重要語群と最も類似している重要語群をもつ表現パターンを用例文データベースから選択し出力することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。さらに、たとえ重要語が誤っていても、その誤りの重度に応じて、正しく文意が理解できる表現に変換するか、意味が理解できなかったことを

ユーザに告知することが可能となり、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0038】

【発明の効果】

以上詳述したように、本発明によれば、表現変換手段は、入力文から重要語を抽出し、抽出された重要語を用いて同じ文意を表す標準的または簡易的表現文に変換し出力することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0039】

また、入力文に含まれる一部の内容語の単語もしくはその変換された単語を重要語として抽出し、この重要語と入力文から推定される文意により決まる表現との組み合わせからなる標準的または簡易的表現文を生成することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。

【0040】

また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0041】

また、入力文から重要語を抽出し、抽出された重要語の共起関係または依存関係から文意を推定し、重要語またはその対訳単語と推定された意図から予め決められている単語のみの組み合わせから標準的または簡易的表現を生成することで、変換規則や文生成規則をコンパクトに構成することが可能となり簡易な処理で表現変換を行うことを可能にする。

【0042】

また、重要語以外の部分で誤った文が入力された場合でも、正しく文意が理解

できる表現に変換でき、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【0043】

また、表現変換手段は、予め、出力される標準的または簡易的表現パターンとそのパターンに当てはまる重要語群及び重要語群間の共起関係または依存関係とを対応させている用例文データベースを作成しておき、入力文から重要語群を抽出し、抽出された重要語群と最も類似している重要語群をもつ表現パターンを用例文データベースから選択し、選択された用例文を出力することにより、上記効果に加え、さらに現実に入力される文の種類やドメインや文型に忠実に確度よく表現変換を行うことを可能にする。

【0044】

また、表現変換手段は、用例文データベースに記載されている表現パターンが重要語または重要語の対訳単語のみで構成されていることにより、さらに強化することが可能である。

【0045】

また、表現変換手段は、用例文データベースに記載されている表現パターン、重要語群及び重要語群間の共起関係または依存関係の重要語の記述を重要語を含む単語クラスの記述とすることで、上記効果に加え、用例文データベースに含まれていない重要語が入力されても妥当な用例文を選択可能となり、より多種多様な入力文に対応できる表現変換が可能となる。

【0046】

また表現変換手段は、入力文から重要語群を抽出し、抽出された重要語間の関係から入力誤りである単語を推定し、推定された単語を除いた残りの重要語から文意を推定し、推定される文意により決まる単語の組み合わせから標準的または簡易的表現を生成することで、上記効果に加えて、たとえ重要語が誤っていても、その誤りの重度に応じて、正しく文意が理解できる表現に変換するか、意味が理解できなかったことをユーザに告知することが可能となり、従来のように誤ったまま表現変換結果を出力するという問題を解決することが可能となる。

【図面の簡単な説明】

【図1】

本発明の実施の形態の表現変換装置を示す図

【図2】

本発明の実施の形態で使用する用例DBを示す図

【図3】

本発明の実施の形態で使用する用例DBを示す図

【図4】

本発明の他の実施の形態の表現変換装置を示す図

【図5】

本発明の他の実施の形態で使用する用例DBを示す図

【図6】

本発明の実施の形態の表現変換装置を示す図

【図7】

本発明の他の実施の形態で使用する用例DBを示す図

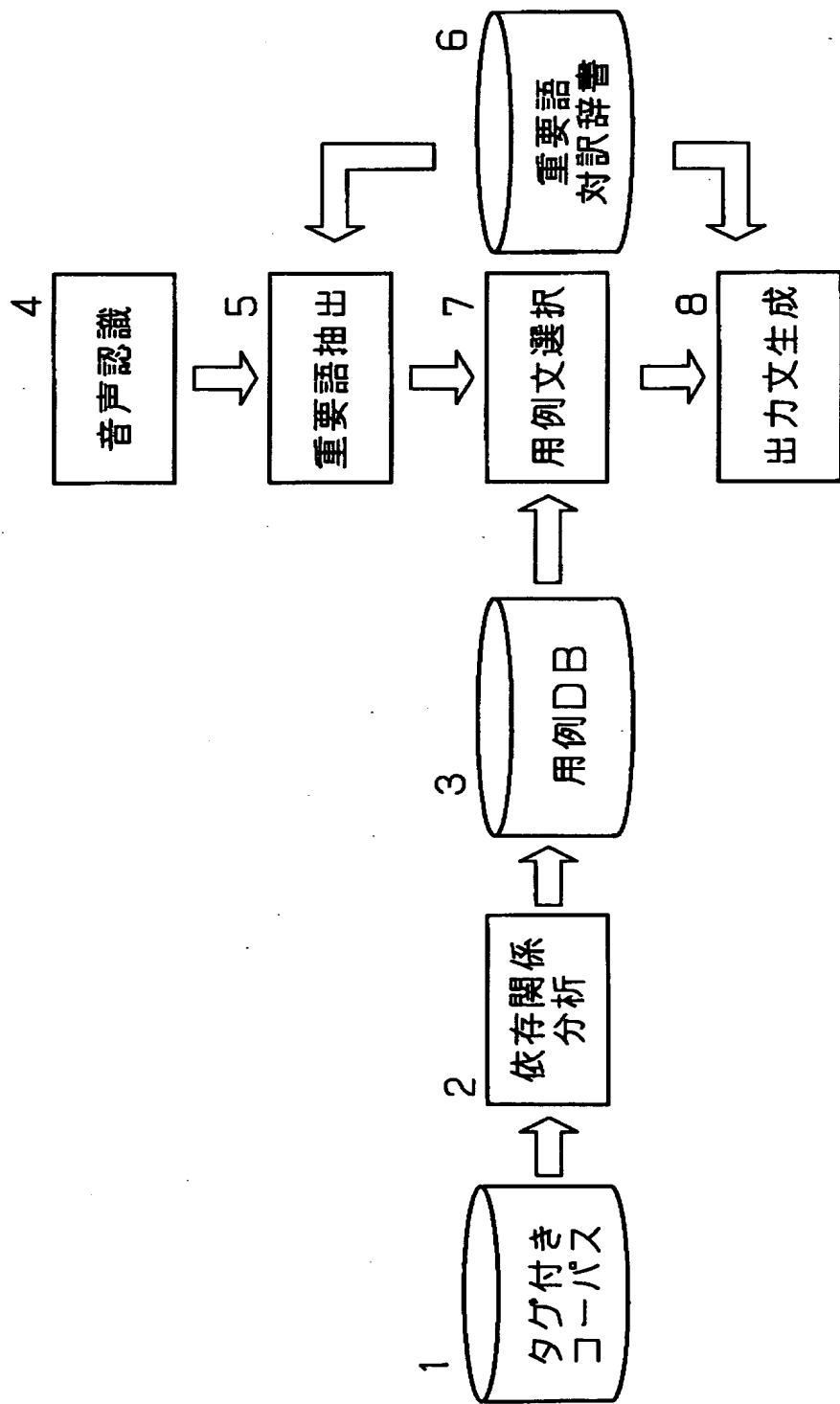
【符号の説明】

- 1 タグ付きコーパス
- 2 依存関係分析
- 3, 11 用例データベース
- 4 音声認識
- 5 重要語抽出
- 6, 16 重要語対訳辞書
- 7, 14 用例文選択
- 8, 15 出力文生成
- 12 分類語彙表
- 13 単語クラス化
- 21 誤認識単語推定

【書類名】

図面

【図1】



【図2】

(a)

重要語対訳辞書例

コーヒー	: coffee
お願い	: *
ミルク	: milk
冷たい	: cold
あり	: *

用例口語例

重要語	依存関係	表現パタン
①コーヒー	② お願い	: (①→②) : I'd like to ② please.
①冷たい	②ミルク ③あり	: (①→②)(②→③) : Do you have a ①②?
...		

(b)

タグ付きコーパス例

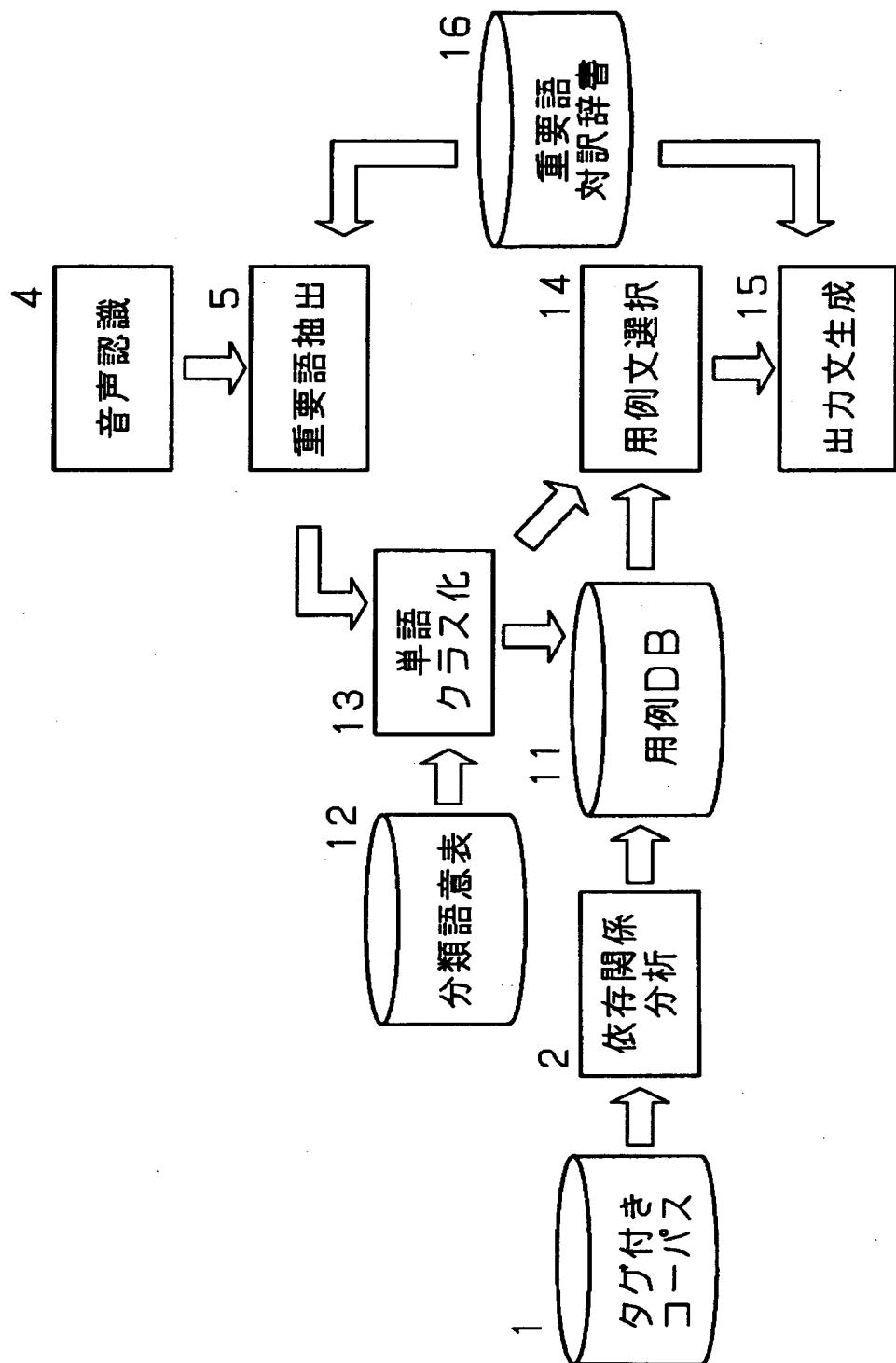
コーヒー (一般名詞) を (格助詞) お (接頭詞) 願い (サ変名詞) し (動詞) ます (助動詞)
冷たい (形容詞) ミルク (普通名詞) は (係助詞) あり (動詞) ます (助動詞) か (終助詞)
...

用例DB例

【図3】

重要語	依存関係	表現パタン
①コーヒー	② お願い	:(①→②) : ②please.
①冷たい	②ミルク ③あり	:(①→②)(②→③) : Any ①②?
...		

【図4】



【図5】

(a)

分類語彙票

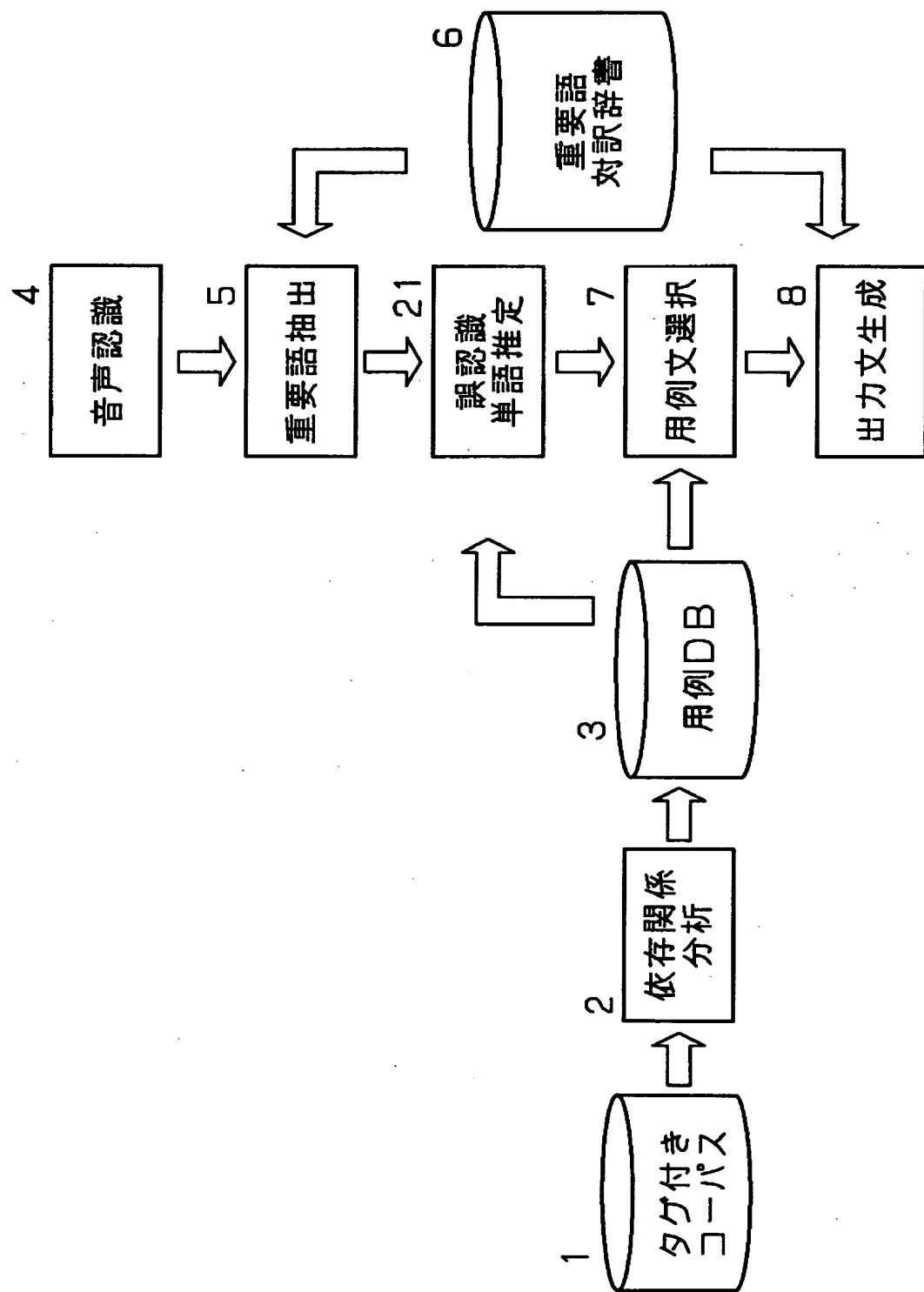
コーヒー (一般名詞)	100
ミルク (一般名詞)	100
紅茶 (一般名詞)	100
...	...
冷たい (形容詞)	200
熱い (形容詞)	200
...	...

(b)

用例口例

重要語	依存関係	用例文
① 100 ② お願い	:(①→②)	: I'd like to ② please.
① 200 ② 100 ③ あり	:(①→②)(②→③)	: Do you have a ①②?
...

【図6】



【図7】

入力文 : 熱いミルクはありますか

認識結果文 : 青いミルクはありますか

用例DB例 :

重要語		依存関係	用例文
①コーヒー	② お願い	(①→②)	② please.
①熱い	②ミルク	③あり (①→②) (②→③)	Any ①②?
...			

抽出された重要語 : 青い、ミルク、あり

上記重要語と用例DBとの依存関係比較結果 :

- (青い、ミルク) ×
- (青い、あり) ○
- (ミルク、あり) ○

【書類名】 要約書

【要約】

【課題】 従来、多数の異なる対訳フレーズとその組み合わせから作成する変換規則が煩雑で膨大になってしまい処理時間もかかってしまう。出力文生成部にて行われる文法的なチェックにも、複雑な規則が必要となり、特にゼロ代名詞の補完などは完全に正しく補完できる技術がなく、時々は誤った補完を行ってしまう。また、音声認識誤りなどにより、部分的に誤っている文が言語翻訳部に入力された場合、誤ったフレーズ内およびフレーズ間規則に基づき言語変換がなさるため、全く意図が伝わらない翻訳結果を出力してしまう。

【解決手段】 入力文中から重要語を抽出し、抽出された単語を用いて入力文意を理解し、重要語と簡易な表現パターンまたは重要語のみを用いて表現変換文を構成することで、言語変換規則と出力文生成規則の簡略化をはかり、コンパクトで高速に処理可能な表現変換手段を提供する。

【選択図】 図1

出願人履歴情報

識別番号 [000005821]

1. 変更年月日 1990年 8月28日

[変更理由] 新規登録

住 所 大阪府門真市大字門真1006番地
氏 名 松下電器産業株式会社